MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2021/2022

DSP5014 – STATISTICS & PROBABILITY

(For diploma students only)

10 AUGUST 2022
9.00 a.m – 11.00 a.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of 6 pages including appendix.
2. Attempt **ALL** questions. The distribution of the marks for each question is given.
3. Please write all your answer in the Answer Booklet provided.
4. Key formulae are given in the Appendix.

QUESTION 1

a) Statistics are frequently being applied in our daily life. Provide one example and explain briefly how statistics applied in it. (2 marks)

b) Identify the population and sample in each following situation. (6 marks)

i. An agent wants to know the opinions of Ayer Keroh residents on the new development of a new housing area. The agent surveys 800 residents.

ii. Coach Ken wants to know how many students in Melaka campus are going to the volleyball game. He surveys 50 students and the result shows ten are going to the game. The rest are not going to the game.

iii. Business Unit of Multimedia International University takes a poll to find out what students prefer to eat during lunch at the campus cafeteria. A random 70 students involve taking part in the poll.

c) A business magazine published data on the best small firms in 2020. These were firms which had been publicly traded for at least a year, have a stock price of at least RM5 per share, and have reported annual revenue between RM5 million and RM1 billion.

i. Table below shows the ages of the CEO for the first 70 ranked firms. Complete the grouped data table. (2 marks)

Age	Frequency	Relative Frequency	Midpoint
$40 \leqslant 50$	14		
$50 \leqslant 60$	29		
$60 \leqslant 70$	16		
$70 \leqslant 80$	11		
	70		

ii. Find the value of variance and standard deviation. (4 marks)

d) The number of people visiting a café each day for 11 days is listed below.

104 131 120 115 109 124 128 118 116 120 125

i. Construct a stem and leaf diagram for the above data. (3 marks)
ii. Find the quartiles. (5 marks)
iii. Are there any potential outliers? (4 marks)

Continued...

e) Mr. Ben recorded the heights (in inches) of his students and represented the data in a stem and leaf diagram as below.

Stem	Leaf							
4	3	4	5	6	7	8	9	9
5	1	3	3	4	4	7	9	
6	0	1	2	2	2	3	4	

Key: 4 | 3 = 43 inches

i. How many students have a height above 60 inches? (1 mark)
ii. What is the mode height of the students? (1 mark)
iii. Find the percentage of students that doesn't have similar heights. (2 marks)

[TOTAL: 30 MARKS]

QUESTION 2

a) Suppose a study of speeding violations and drivers who use cell phones produced the following fictional data:

	Speed violations (S)	No speed violations ($\bar{S}$)	Total
Use cell phone while driving (C)	25	280	305
Does not use cell phone while driving ($\bar{C}$)	45	405	450
Total	70	685	755

Calculate the following probabilities using the table.

i. $P(\bar{S})$ (2 marks)
ii. $P(C \cup \bar{S})$ (3 marks)
iii. $P(\bar{C}|S)$ (3 marks)
iv. Are $P(C)$ and $P(S)$ mutually exclusive? (1 mark)

Continued...

b) Two files are drawn without replacement from a shelf containing five red and three green files. Find the probability that

i. both files will be red. (3 marks)

ii. both files will the same color. (4 marks)

iii. at least one file will be green. (4 marks)

[TOTAL: 20 MARKS]

QUESTION 3

a) Joyce had just begun her new job as a sales executive of a very competitive company. In a sample of 36 sales calls it was found that she closed the contract for an average value of RM108 with a standard deviation of RM12. Test at 1% significance level that the population mean is at least RM100. (12 marks)

b) To investigate a possible relation between gender and handedness, a random sample of 320 adults was taken, with the following results:

	Men	Women
Sample size, n	168	152
Number of left-handed, x	24	9

Test at 5% significance level that the proportion of men who are left-handed is greater than the proportion of women. (13 marks)

Note: Provide your proportion in 4 decimal places.

[TOTAL: 25 MARKS]

Continued...

QUESTION 4

A lecturer would like to examine the relationship (if any) between the number of homework questions her students do during the semester and their final grade. She randomly selects 12 students, and the data are presented as follows.

Number of Questions	Final Grade
51	62
58	68
62	66
65	66
68	67
76	72
77	73
78	72
78	78
84	73
85	76
91	95

Note: Provide your answer in 2 decimal places.

a) Calculate SS_{xx}, SS_{yy} and SS_{xy} based on the information given. (9 marks)

b) Calculate the correlation coefficient, r and interpret the answer. (4 marks)

c) Find the linear regression equation $\hat{y} = \beta_0 + \beta_1 x$. (5 marks)

d) Predict the number of questions the student does if his/her grade is 86. (2 marks)

e) Compute the value of *SST*, *SSR* and *SSE*. (3 marks)

f) Calculate the standard error of the estimate, s_e. (2 marks)

[TOTAL: 25 MARKS]

End of Paper

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{\left(\sum f_i\right)-1}\left[\sum m_i^{\,2} f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}\right]$ where m : class midpoint f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}}$	$Z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\hat{p}_p\left(1-\hat{p}_p\right)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ *where* $\hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$
- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{(S_{xy})^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error,	Regression t-Test	Correlation t-Test
$s_e = \sqrt{\dfrac{SSE}{n-2}}$	$t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	$t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$